GLOBAL GENE EXPRESSION ANALYSIS TO UNAMBIGUOUSLY IDENTIFY HOST GENE RESPONSES CHARACTERISTIC OF EXPOSRE TO BIOTHREAT AGENTS

Rasha Hammamieh* and Marti Jett
Molecular Pathology, Walter Reed Army Institute of Research, Silver Spring, MD 20910

ABSTRACT

We are studying the complex interaction between various biological pathogens and the host to understand the basis infectious or biothreat-induced diseases and to identify host defense strategies and the mechanisms by which they are regulated. Although gene response profiles show unique signatures quite rapidly after exposure, they also have the potential to reveal phases of progression of illness to a) provide stage-specific diagnosis and b) identification of potential molecular targets for stage-appropriate therapeutic interventions for intractable illness induced by unconventional pathogenic agents.

For this approach, several issues required prompt solutions including a) establishment of a baseline for "normal & healthy" individuals b) ability to fill in the gaps inherent in vivo studies with in vitro findings c) differentiating biothreat induced flu-like illness from flu or other common illness d) harnessing the power of prior knowledge to correlate with the global gene responses, e) as well as certain other factors.

We have used a library of 20,000 human cDNA (~10,000 are known genes) to construct customized microarray chips used in these studies. We determined gene expression in human peripheral blood mononuclear cells (PBMC) in response to 15 pathogens at different time points in vitro (3-5 replicates). This provided a framework for us to then utilize responses in animal models that closely imitate the illness as it occurs in humans. For those studies, PBMC or whole blood were collected at various time points post exposure to track the primary, secondary and subsequent gene responses elicited by the pathogenic agents. The massive amounts of data are overwhelming but provide an incredibly rich source for both diagnostic and therapeutic approaches.

The scientific community has realized the potential of these massive studies. Clever, far-reaching data mining approaches have been devised which we have utilized. Of necessity, we developed and customized certain software ourselves including a MIAME compliant relational database that integrates with external databases such as PubMed, LocusLink, GeneCard, Hugo gene ontology database and Biocarta and KEGG pathway databases. The links are invaluable in data mining and evaluating host response to various pathogens. We have also developed a word-search clustering software that automatically searches PubMed for up to 200 genes at a time seeking documentation of physiologic function to explain stage-specific clinical/pathological observations.

This information is aimed at diagnosis, predicting the course of impending illness and identifying appropriate therapeutic targets at different stages. A most critical aspect is to minimize interpretation difficulties by establishing pathogen-specific signatures that can be readily distinguished from "normal /healthy baseline" profiles or common illnesses with similar initial symptoms. Therefore, we analyzed data (obtained over ~4 years) from 75 healthy donor "control" samples of different ethnicity, sex and age range of 18-36 years.

Microarray gene expression data were analyzed for the control samples to create a base line for gene expression to be used in our studies. For this purpose, we especially focused on genes that were expressed at approximately baseline levels (barely detectable) in the 75 control samples and exhibited high expression upon exposure to at least one pathogen. Out of these low-expression genes in samples from healthy controls, we identified those that became overexpressed upon exposure to various pathogens. From these genes, pathogen-unique patterns were found, even at early time points. We are evaluating devices that permit rapid hybridization/testing on inexpensive platforms that could be used for wide-spread screening in event of suspected exposure to unconventional pathogens to differentiate from common infectious illnesses.

We have identified host gene expression patterns that can discriminate exposure to various biological threat agents. Each of these gene patterns regulated by a specific agent reveals the cascade of events that occurs after the host encounters a pathogenic agent. Even though these pathogens initially cause similar symptoms, such as malaise, fever, headache, and cough, the course of illness induced by each of them differs in time frame of illness patterns. Using these signature gene profiles to assess possible exposure to pathogenic agents or to differentiate them from non lethal illnesses when the classical identification of a pathogen is not conclusive may fill a gap in the arsenal of diagnostic tools. Rapid detection, before the symptoms appear or even at various stages of illness, offers the opportunity to initiate appropriate treatment. Furthermore, this technique may provide the means to identify new therapeutic approaches to ameliorate the devastating results of these pathogens.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 DEC 2004	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Global Gene Expression Analysis To Unambiguously Identify Host Gene Responses Characteristic Of Exposre To Biothreat Agents	**g**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Molecular Pathology, Walter Reed Army Institute of Research, Silver Spring, MD 20910**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001736, Proceedings for the Army Science Conference (24th) Held on 29 November - 2 December 2005 in Orlando, Florida., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**8**	

1. INTRODUCTION

Preparations for the Army of the future include reliance on high technology instrumentation for diagnostic and therapeutic approaches. This highly sophisticated hardware requires extensive computational capabilities and will have the potential to provide a wealth of important information to assist in keeping the warfighter healthy and ready for action. The physical circumstances that can exist in various theaters of combat could result in exposure of the warfighter to unusual endemic pathogenic agents or environmental toxins not previously encountered. In addition, deliberate biological threat exposure must be differentiated from illness induced by common pathogenic agents.

For the Army of the future, various devices are under development for real-time determination of easily measurable vital signs and other clinical parameters. Due to the above-mentioned hazards in remote places where troops may need to carry out their mission, rapid determination could be critical to differentiate the urgent medical condition that would result due to biothreat exposure vs common flu-like illness or endemic non-lethal pathogen. We have a vision to address this scenario and the eventual aim would be to utilize host gene expression responses to biothreat pathogenic agents to differentiate them from common flu-like illnesses. The advantage of relying on host gene expression rather than direct pathogen identification, is that in a few drops of blood (sufficient for gene analysis) are hundreds of thousands of lymphocytes that have coursed through even remote areas of the body (lungs, lymph nodes, liver, etc) searching for "invaders". During this reconnaissance role, when these cells find a pathogen they react to neutralize it, creating a record (unique to each pathogen) of the encounter in their messenger(m) RNA. As a result, the host gene expression response can be determined very early, or even at any time post-exposure. In fact, for exposure to one biothreat toxin, a unique signature is observed by at least 30 min post-exposure in non-human primates (NHP), yet the initial onset of illness did not occur until 4 h post exposure. For one bacterial infection, Actinobacillus pleuropneumoniae in swine, we have seen unique gene signatures with in 2 h although onset of even mild malaise did not occur until ~10 h. However, it is not just very early detection that can be carried out using host gene expression responses, for we observe that there are stage-specific host responses that can define the "course of impending illness".

We are currently creating a library of host gene expression responses to biothreat and certain common pathogenic agents. This process utilizes the massive gene chips that interrogate 20,000 (cDNA) or 40,0000 (oligonucleotides) genes. However, our eventual plan is to select sets of genes that can be used as signatures from the library of host responses and proceed to utilize small "macroarray" chips containing these carefully selected sets of genes that would differentiate among many common vs biothreat pathogenic agents. Many commercial efforts are underway to construct such small devices (even hand-held instruments) that can directly use the small sample without derivatization, and utilizes technology for "instant hybridization". Some current approaches even use RNA directly rather than conversion to cDNA, as is the usual custom for microarrays. In general, the technology that is currently under development for such devices offers potential for revolutionary measures to use for not just detection of exposure to pathogenic agents but also to design treatment regimens that are tailored to the stage of advancement of the illness and can meet needs of the individual warfighter.

2. EXPERIMENTAL APPROACH

In these studies, we are creating a library of gene expression responses in peripheral blood mononuclear cells (from exposures in vitro and for some pathogens, in vivo as well) to anthrax, brucella, dengue, cholera, plague, staphylococcal enterotoxins (SE), and other biological threat and common pathogenic agents using up to 20,000 cDNA gene microarrays. The cDNAs are maintained by us and commercially printed onto microscope slides at 10,000 genes per slide.

2.1 Description of the system.

Our system permits 2-color competitive hybridization and we have utilized that by comparing all samples to a "universal reference RNA standard" (Strategene). In essence, the reference RNA is flluorescently labeled with Cy 3 and separately the sample from pathogen exposure is labeled with Cy 5 (and visa versa). In this way, every sample used is compared to the exact same RNA; that has the advantage to normalize the inevitable variations that occur from year to year, with different personnel carrying out the techniques and variations that may occur among batches of microarray chips. Experiments were carried out in replicates at each time point for each pathogen using the cDNA microarrays.

2.2 Initial Image acquision and data processing.

Images of the array slides are acquired and processed to produce a data file that contains thousands of values for each experiment. We have used Axon's GenePix scanner and software for microarray data visualization

and interpretation. Results were then confirmed using real time Rt-PCR.

We used the reference design where a reference RNA sample is co-hybridized with each sample on the slide. This design allows us to normalize between the slide for variations that can be due to hybridization, transcription and labeling efficiencies (technical variations). We used various modules to analyze the microarray data including GeneSpring, Partek Pro, SAM and Bioconductor. Using Analysis of Variance (ANOVA) we determined genes that exhibited variations in expression between the control samples. These variations may be due to many factors including biological and technical variations. These normally varying genes are excluded from further analysis to study gene regulation upon exposure to pathogens. GeneSpring microarray data analysis software was used for data analysis, gene clustering, studying patterns of gene expression and exploration of pathways altered by each pathogen.

.

2.3 "Project Normal" for Healthy Humans

We created a base line for gene expression in PBMC obtained from 75 healthy donor "control" samples of different ethnicity (African American> Hispanic . Caucasian >> Asian descent), sex and age range of 18-36 years. We analyzed gene expression data for the control samples to identify genes that were normally varying among healthy humans of diverse ethnicity. These genes were excluded from further analysis since their expression was so inconstant among these individuals. We were interested in finding genes that can be used as markers for an exposure in the case of an outbreak where controls are hard to identify.

2.4 Minimizing ambiguity: Selection of off/on genes.

We selected genes that were expressed at near baseline levels (barely detectable) in the 75 control samples and were highly expressed upon exposure to at least one pathogen (off-on regulated genes). Out of these genes that were expressed near the baseline levels in all control samples and were shown to be highly expressed upon exposure to various pathogens, sets of genes were unique for certain pathogens at early time points. Conversely, we also determined genes that were highly expressed in all the control samples and were barely detectable upon exposure to a certain pathogen. We confirmed our results using real time-PCR. These genes have the potential to be diagnostic markers for exposure to a specific pathogenic agent.

2.5 Development of new techniques for data mining

We also developed a word search and clustering software called GeneCite to do multiple queries searching the PubMed literature database. This program provide PubMed search for 200 genes at a time and gives a score to gene relatedness in the literature. This software provides a fast tool for data mining and gene regulation studies.

Another tool we developed is called PathwayScreen and is used to screen a list of genes of interest against a pathway database to resolve pathway regulated by certain treatment. This tool offers a fast and high throughput pathway analysis for microarray data.

3. RESULTS AND DISCUSSION

When we evaluate various biological warfare pathogens at different time point, the massive amount of data is overwhelming and it is a very important source for both diagnostic and therapeutic approaches.

We have utilized many software packages and developed some of our own as well for microarray data evaluation. We have developed a relational database software package that tracks all information required about each sample and experiment. Using this relational database, we are able to get information about alteration in expression of genes of interest or a specific pathway by one or more pathogen and to find genes unique for each pathogen. Furthermore, this database is linked to external databases such as PubMed, LocusLink, GeneCard, Hugo gene ontology database and Biocarta and KEGG pathway databases.

We have applied various clustering techniques to group genes with similar expression patterns or functions. Most cluster analysis methods are hierarchical; the resultant classification has an increasing number of nested classes and the result resembles a phylogenetic classification. Non-hierarchical clustering analyses are also used, such as K-means clustering and self-organizing method (SOM), which partition genes into different clusters without specifying the relationship between individual elements.

3.1 Project Normal: Baseline gene expression

This information is aimed at diagnosis, predicting the course of impending illness and identifying appropriate therapeutic targets at different stages. A most critical aspect is to minimize interpretation difficulties by establishing pathogen-specific signatures that can be readily distinguished from "normal /healthy baseline" profiles. Therefore, we analyzed data (obtained over ~2 years) from 75 healthy donor "control" samples of different ethnicity, sex and age range of 18-36 years. We found that <10% of the total number of the genes on the arrays exhibited variation in expression between the

slides. These genes were eliminated from further analysis for genes regulated upon exposure to the pathogens (Fig.1).

Figure 1.

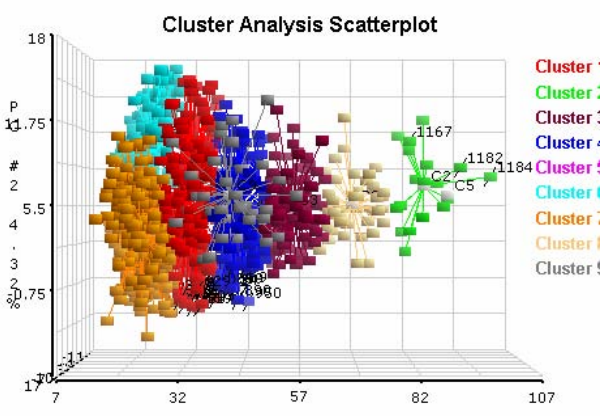


Figure 1. Three dimensional scatter plot for the first three principal components of a PCA analysis. The genes represented by the clusters at the left side of the graph, were expressed universally among the 75 people who comprised various groups of diverse, healthy individuals. Genes that are normally varying between the slides are plotted in green (right side of graph).

3.2 Host responses upon pathogen exposure

Exposure to pathogen was carried out using parallel PBMC from the healthy donors (Fig 1) in which one group of samples were exposed to the pathogen and the other used as a control. Relative to the control (harvested at the same time as the "post-exposure" sample) the changes in each gene was catalogued and the dendrogram (Figure 2) constructed. There were at least 3 replicates of each exposure and the aim was that these be from different people in order to identify any gene responses that could be due to some unique aspect of one specific individual. For preparation of the dendrogram, the only consistent changes among the individuals was recorded.

3.3 Minimizing ambiguity by selection of off/on genes

We especially focused on genes that were expressed at baseline (barely detectable) levels in the 75 control samples and overexpressed upon exposure to at least one pathogen. There are many fascinating scenarios, such as highly expressed CD markers in control cells that simply disappear upon pathogen exposure. Some of this may be due to sequestration of subsets of cells, a phenomenon that has previously been described as a "signature" for certain of the biothreat toxins (SEB, for example). Figure 3 shows the genes, relative to Figure 2 that are essentially turned "ON" upon exposure to each of the pathogenic agents.

Figure 2.

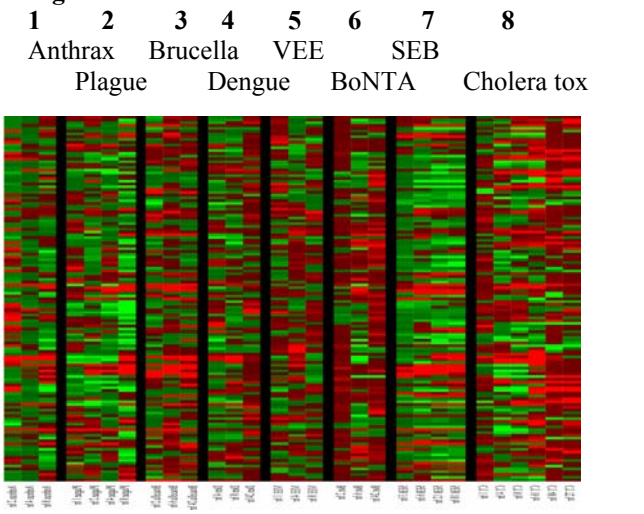
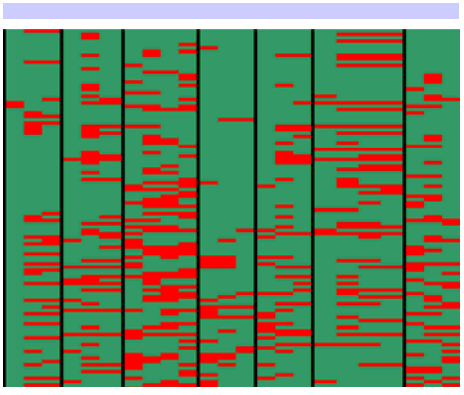


Figure 2. A pseudo-color cluster analysis of genes regulated by eight different pathogens at various time points. The figure is arranged to show the changes induced by B. anthracis exposure (Anthrax-far left, #1) at 3 different times of exposure (up to first black line). Plague exposure was carried out at 4 different time periods (group #2), and so on as indicated at the top of the graph (through #8, Cholera toxin). Increased (red) or decreased (green) gene expression is illustrated for each pathogenic agent at from 3-5 different time lengths of exposure

Figure 3. Genes that are expressed below the background levels in the control untreated samples and are up regulated by one or more pathogens. That is, these genes were hardly detectable in control cells from



diverse donors, but upon exposure to pathogens, these genes became massively overexpressed and the change in expression levels would be clear and could be part of an algorithm for eventual future use. Similarly, we wanted to identify genes that were expressed at reasonably high levels in the normal healthy individuals, but were turned "OFF" upon exposure to biothreat

agents. CD markers on lymphocyte subsets is a good example of this scenario and may indicate sequestration of certain subsets of cells..

3.4 Demonstration of specificity of the 'ON/OFF' genes

We used Real time-PCR to validate this approach for some of the genes that were turned 'ON' upon exposure to a pathogen (Figure 4 a-c). By selecting some of these specific genes from the cDNA microarrays and re-evaluating them using real time PCR, we were able to identify genes that were barely detectable in **all** the control samples (very low copy number) and were highly expressed when cells were exposed to a pathogen. These particular genes were unique for certain pathogens and were expressed only when cells were treated by that pathogen. No expression was detected for these genes when cells were exposed to other pathogens. We anticipate that ~10-20 such genes would be needed to completely identify each pathogen.

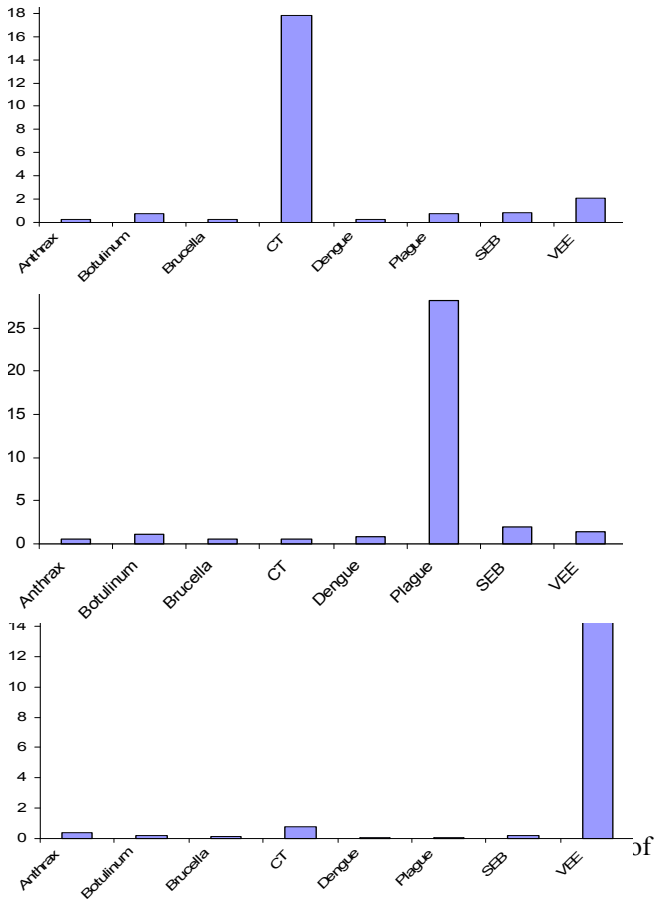


Figure 4. Example of genes turned 'OFF' in control samples that become massively turned 'ON' upon exposure to pathogen. Real time-PCR of genes that were solely expressed in PBMC treated with the cholera toxin

3.5 Functional Genomics: Data Mining/Mechanisms

We developed tools for data mining and biological interpretation of the meta data. One of the tools, PathwayScreen, provides a high throughput pathway analysis for genes regulated by certain pathogen. It captures a list of LocusLink ID numbers for the genes of interest and outputs a file listing the pathways that those genes are in and a link to any appropriate pathway database, namely BioCarta.com. or KEG.

Pathway Name	A few specific key Genes in this pathway	Locus Link ID
Dentritic Cell	Intercellular Adhesion Molecule 1	3383
	CD8 Antigen	925
	alpha Polypeptide (pCD4 antigen)	920
T Cell Surface Markers	Lymphocyte-sp Protein tyr kinase	3932
CTL Mediated Apoptosis	Intercellular Adhesion molecule1	3383
	CD8 Antigen	925
	Gamma Polypeptide Antigen	917
CTL Surface Molecules	T cell receptor β locus	6957
	T cell receptor α locus	6955
T Cell Co-stimulatory signal	Ls-specific protein Tyr kinase	3932
	Zeta chain TCR protein	7535
	Ls-specific protein Tyr kinase	3932
Cell Signalling Pathway	PKC β1 locus	5579
	PKC α	5578
	Mitogen activated Prot Kinase 3	5595
Cell Transcrip Factors	alpha Polypeptide (pCD4 antigen)	920
	Gamma Polypeptide Antigen	917
	CD8 Antigen	925
Many other pathways have been defined in BioCarta & KEGG		

Figure 5a. When PathwayScreen is applied a list of genes, a tab delimited text file report is created and can be accessed using any spreadsheet program. This file contains the names of the pathways the genes are included in, the url where these pathways can be viewed, and the genes from the original list that are in the pathways – both the gene names and the gene Locus ID numbers are included as shown in the examples above

When the list of gene showing changes is established for each study, they are imported into Pathway Screen and each gene (about which functional details is known) is assigned a mechanistic pathway using the BioCarta database. Gene Spring has a similar output relating to the metabolic pathways in KEGG. BioCarta pathways usually relate to biochemical rather than strictly metabolic pathway cascades. An example of such a BioCarta pathway (Figure 5b) details the regulatory mechanisms for Angiotensin Converting Enzyme (ACE-1). The corresponding RNA for ACE-1 was remarkably upregulated for in vivo studies of lethal shock induced by

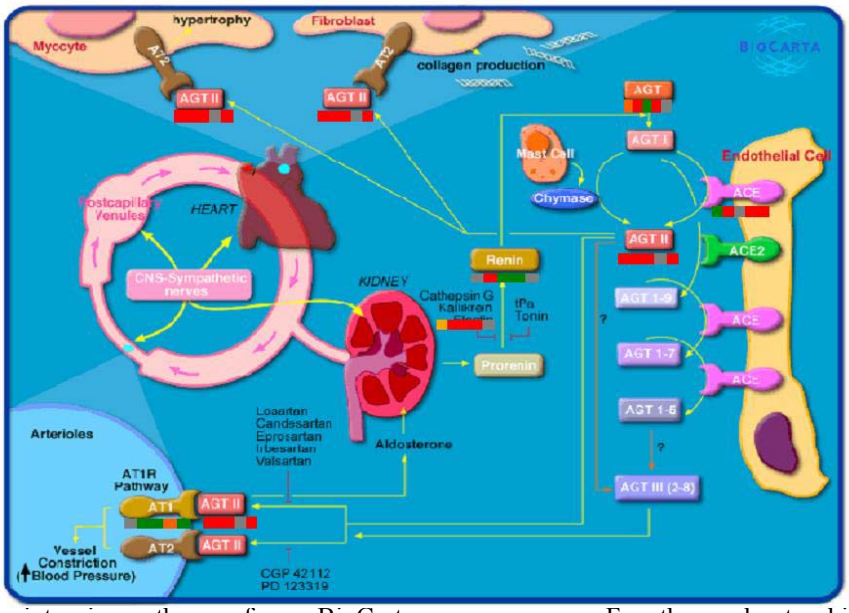


Figure 5b. Angiotensin pathway from BioCarta containing the gene responses observed upon challenge of piglets with SEB. Near the top left corner is the Angiotensin 2 receptor (AT2) interacting with Angiotension II (AGT II) and under it is a "bar" that consists of 5 segments (1 segment for each time period). For AGTII, the first 3 time periods (2, 6, 24h) show upregulation (red), data for the 4th time period (48 h) is missing (grey) and the last segment (72 h) is also upregulated (red). Each major component of the pathway has these segmented "bars" under the name of the mediator. As illustrated in the lower left corner, the eventual action is blood vessel constriction. This is a major problem for lethal shock, since at one point, attempts to increase blood pressure result in hemorrhage into the tissues leading to multi-organ failure and death. Establishment of biochemical pathways such as this provides a frame of reference to use for designing new therapeutic strategies for specific stages of the illness.

The reason this pathway was of interest to us relates to our observations of the genes showing altered regulation prior to the onset of severe vascular leakage in a model of lethal shock induced by SEB (Figure 5b).
Another tool we developed, GeneCite, offers a high throughput query of the PubMed database for citations using search terms taken from an input file (i.e. the list of genes). Due to the limitations of the Excel spreadsheet, just 200 genes of interest can be searched simultaneously. The output file is a spreadsheet with the gene names in the first column and the number of citations is in the next column (Fig.6). There are three ways we can use GeneCite. a) The first is a simple unrestricted search of the literature to see what may be known about each specific protein.

For those about which little is known, this approach could be useful. b) The second use is to attempt to sort based on function and for that approach, we have developed lists of clinical descriptions related to the course of illness induced by the pathogen. For example, several biothreat agents eventually produce devastating effects by leading to lethal shock. Therefore, we developed a list of 65 search terms related to lethal shock, such as the 4 terms shown in Fig. 6, columns 2-5. Some of the other terms in that search strategy include capillary dilation, fibrin, DIC, ischemia, vascular leakage, etc. As is shown in Figure 6, the number of "hits" is recorded and a mouse click on that number brings up the list of publication titles, abstracts and PubMed links in which each gene/protein of interest has been previously characterized in relation to the search term. In cases of many hits, that information may be intuitively known but for those genes for which there are few publications, this has already helped immeasurably to begin to correlate alteration in gene expression (with a clue as to the protein's function) along a time line related to clinical manifestations of the illness. Functional genomics approaches provide incredibly rich information that can potentially produce diagnostic markers of impending illness at a time frame early enough to initiate appropriate treatment. For unidentifiable pathogens (natural or deliberately altered pathogens) it could be possible to track the functional characteristics of host responses in order to predict onset of clinical manifestations. Clearly, new therapeutic targets could also be identified. c) Another use is to search for a list of genes against itself (200 x the same 200 genes) in order to discover correlations that are not well-know or well characterized. This has provided a multitude of new information that was obscured in the literature and has helped to expand the biochemical pathways now described in BioCarta and KEGG.

A few genes from a *GeneCite* search	Terms defining aspects of lethal shock (4 of 65 used for searches)			
GENES	**Edema**	**Micro-emboli**	**Infiltra-tion**	**Lethal shock**
Cytokine inducible SH-2	0	0	0	0
Lymphotoxin Beta	1	0	9	0
Lymphotoxin β Receptor	1	0	5	0
Protein Regulator of cytokinesis 1	0	3	3	1
LPS-induced TNF-α factor	10	7	48	65
TNF receptor-associated factor -1	0	0	1	0
Janus kinase 1	0	0	1	0
bradykinin receptor B2	47	0	10	0
PI3 Kinase	6	1	3	0
Ubiquitin associated prot	3	0	6	2
Phospholipase A2	263	54	67	8

Figure 6. A small portion of a screen shot of the output files produced after applying GeneCite in which lists of up to 200 genes were searched against 65 terms relating to aspects of lethal shock. .

3.6 Supplementing *in vivo data with in vitro studies*

In vitro studies provide a potential wealth of information, but eventually we need *in vivo* confirmation of experimental findings. In our current studies, we have amassed data to differentiate host gene expression responses among numerous biological threat agents using peripheral blood mononuclear cells (PBMCs) to create a record of exposure to pathogenic agents. We first carried out those studies *in vitro*, exposing human PBMCs from healthy blood donors to various biological threat agents and analyzing the gene expression changes elicited by the threat agent using cDNA microarray technology. We then confirmed the gene expression patterns, analyzing PBMCs from NHP exposed to *B. anthracis*, SEB or other pathogens (Das et al., 2002; Das et al., 2003). However, one must consider the need to characterize the effects of exposure variables including different doses and exposure times, such that the demands on the use of NHPs are impractical, thus necessitating the exploration of an alternative animal model.

To pursue studies on therapeutic intervention in SEB intoxication, an ideal animal model would express the same pathologic symptoms and responses (e.g., emesis, diarrhea, hyperthermia, shock, neurobehavioral symptoms, death) as humans/monkeys to SEB intoxication at reasonably comparable doses, but also be relatively inexpensive, easy to handle and manipulate, and have specific reagents available for molecular analysis (Jett et al., 2001). Although the spectrum of response of monkeys to SEB is similar to humans, they are expensive and difficult to handle, compromising experimental design, and measurement. On the other hand, mice are easy to handle and cheap, however, three models based on mice have the disadvantage of requiring pre-sensitization, and the spectrum of response in the mouse models is not the same as humans.

We have developed a model of SEB-induced lethal shock using piglets. Piglets are also reasonably inexpensive and the experiments require simple housing for short intervals during the experiment. They are locally available and are routinely delivered from a USDA approved facility. Swine models for hypovolemic shock and other cardiovascular disorders have been well-studied for decades**.**

Our studies with SEB-induced lethal shock in piglets shows that their clinical responses and pathology closely correlate with those same parameters as characterized in NHP models (Mattix et al., 1995).

Although the use of DNA microarray technology for the study of gene expression in piglet tissues is certainly informative, several concerns are apparent that do not exist for tissue cultures. Even genetically identical organisms housed under the same conditions are likely to have a different hormonal milieu. The state of the immune system and the degree of inflammatory activity can cause global changes in gene expression from piglet to another. This is mostly problematic in studies involving toxic shock or stress responses. We have determined gene expession profiles in normal healthy piglets to establish a baseline (Hammamieh et al., 2003).

Microarray data from both *in vitro* and *in vivo* conditions in piglets were analyzed and genes were clustered to show patterns of expression. We applied ANOVA to determine genes that show differences in expression patterns between *in vitro* and *in vivo* experiments with a p-value <0.05.

Principal component analysis was conducted using these genes, showing that the *in vivo* and *in vitro* conditions are distinguishable when this set of genes was used (Fig. 7a).

We applied a class prediction method where the algorithm learns from gene expression patterns in SEB in vitro training set. This algorithm determines best predictor genes using the training set which can be used to predict test samples using the k-nearest neighbor algorithm. We then examined how well the algorithm discriminated SEB among other toxin treatments in the test data set that was composed of SEB in vitro, SEB in vivo, cholera toxin, and botulinum toxin. We were ableto identify a subset of genes that correctly predicted 5 out of 7 in vivo SEB treatments to be SEB when compared to other toxins. We applied Principal component analysis using this subset of genes; Figure 7b shows no distinguishable difference between the profiles.

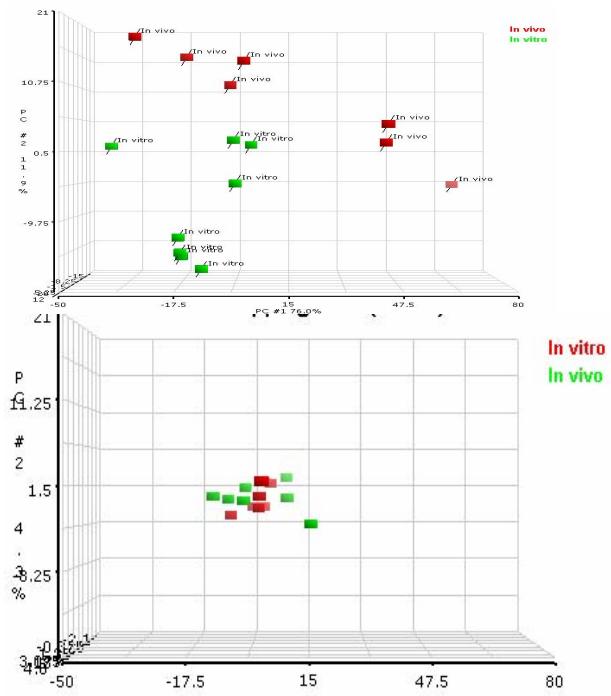


Figure 7a. Principal component analysis for genes differentially expressed between *in vivo* and *in vitro*. ANOVA was carried out to identify genes that exhibited differences in expression between *in vivo* and *in vitro* upon exposure to SEB.

Figure 7b. Principal component analysis for genes that showed similar expression patterns between *in vivo* and *in vitro*. This list was obtained by eliminating the genes that exhibited differential expression between *in vitro* and *in vivo*. Thus, these data indicate that pathogen profiles derived from expression analysis of less than 1200 genes, regardless of the *in vitro* or *in vivo* source of data, can be used to discriminate SEB from other pathogens.

4. CONCLUSION

State of the art biotechnology approaches require serious issues to be addressed in management of massive datasets that are produced in the course of the studies, as well as analysis and mining of the information. Our laboratory has focused of utilization and modification of existing software as well as development of specific software to aid in data mining efforts and other specific needs. Development of predictive mathematical modeling simulations to relate bioinformatics findings with courses of illness progression in lethal shock offers important opportunities for data mining, but primarily provides a framework whereby projections for multiple parameters can be made for many biological threat agents.

We have identified host gene expression patterns that can discriminate exposure to various biological threat agents. Each of these gene patterns regulated by a specific agent reveals the cascade of events that occurs after the host encounters a pathogenic agent. Even though these pathogens initially cause similar symptoms, such as malaise, fever, headache, and cough, the course of illness induced by each of them differs in time frame of illness patterns. Using these signature gene profiles to assess possible exposure to pathogenic agents or to differentiate them from non-lethal illnesses when the classical identification of a pathogen is not conclusive may fill a gap in the arsenal of diagnostic tools.

In the case of an outbreak, it is not easy to identify uninfected control patients. It is very important to identify markers that are signature for each pathogen without the need for a control to normalize to. We identify genes that are not expressed in the base line and are expressed at high levels in treated cells. Using high throughput gene expression analysis along with the proper classification and feature selection algorithms we are able to determine signatures for some of the biological threat agents that can be used to develop a diagnostic tool for these agents. Rapid detection, before the symptoms appear or even at various stages of illness, offers the opportunity to initiate appropriate treatment. Furthermore, this technique may provide the means to identify new therapeutic approaches to ameliorate the devastating results of these pathogens.

5. REFERENCES

1- Das, R., Cummings, C., Mendis, C., Neill, R., Hoover, D., Lindler, L., Paranavitana, C., Huang, X., Ludwig, G., Henchal, E., Yang, D.C.H., Jett, M., 2002. Identification of changes in gene expression induced by toxic agents: Implications for therapy and rapid diagnosis. In: Wade, J. (Ed), NATO Conference: Operational Issues in Chemical and Biological Defense Human Factors in Medicine Panel. RTO-MP-075-019.

2- Das, R., Neill, R., Hammamieh, R., Ludwig, G., Jett, M., 2003. Host gene rxpression profiles in peripheral blood mononuclear cells: Detection of exposure to biological threat agents. ASM Biodefense Research Conference, March 9-12, 2003, Baltimore, MD, USA.

3- Hammamieh, R., Bi, S., Mani, S., Chakraborty, N., Mendis, C., Das, R., Jett, M., 2003. Genetic variations in peripheral blood mononuclear cells in piglets used as an animal model for staphylococcal enterotoxin exposures. Omics. 7(4), 401-409.

4- Henchal E. A., Teska J. D.,. Ludwig G. V,. Shoemaker D. R and . Ezzell J. W, *Clin Lab Med.* **21**, 661-678 (2001).

5- Jett M. *et al.*, Identification Of Changes In Gene Expression Induced By Toxic Agents: Implications for Therapy and Rapid Diagnosis, J. Wade, Ed., NATO: Portugal (2001).

6- Jett M. *et al.*, *Infect Immun* **62**, 3408-15 (1994).

7- Mattix, M.E., Hunt, R.E., Wilhelmsen, C.L., Johnson, A.J., Baze, W.B., 1995. Aerosolized staphylococcal enterotoxin B-induced pulmonary lesions in rhesus monkeys (Macaca mulatta). Toxicol Pathol 23(3), 262-268.